DOCUMENT RESUME

ED 459 473 CS 217 779

AUTHOR Miller-Whitehead, Marie

TITLE A Review of the Literature on Early Childhood Cognitive Academic Language Assessments Suitable for Use in Ethnically Diverse School Systems.

PUB DATE 2001-11-14

NOTE 32p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (30th, Little Rock, AR, November 14-16, 2001).

PUB TYPE Information Analyses (070)

EDRS PRICE EDRS Price MF01 Plus Postage. PC Not Available from EDRS.

DESCRIPTORS *Diagnostic Tests; Early Childhood Education; Evaluation Methods; *Language Tests; *Limited English Speaking; Literature Reviews; Second Language Learning; *Student Evaluation; *Test Selection; Testing; *Validity

ABSTRACT

Although there are many regions of the country that have historically been ethnically and linguistically diverse, school system data as well as information from the 2000 census and other sources has indicated that many hitherto homogeneous regions of the country and their educational institutions have experienced significant growth in populations of immigrant children and their families. Thus, many school systems have faced new challenges to meet the needs of these children from diverse cultural and linguistic backgrounds, including issues related to providing bilingual or English-as-a-Second-Language (ESL) education for their Limited English Proficient (LEP) children. Sources were selected from current extant psychometric reviews, the U.S. Department of Education, the Council of Chief State School Officers, the Center for Applied Linguistics, the various ERIC Clearinghouses, the U.S. Census Bureau, and position and policy papers on ESL and English Language Learning student testing and placement from the American Educational Research Association (AERA), the National Center for Research on Evaluation, Standards, and Student Testing (CRESST), and the Center for Research on Education, Diversity & Excellence (CREDE). The review examined the literature for approaches used to determine content, construct, and predictive validity of kindergarten cognitive diagnostic assessments and their usefulness for making decisions about prekindergarten students and groups of students from ethnically diverse populations. The paper: (1) identifies the most current literature related to the examination of item domain goals and objectives, sources for sample items that could be useful to equate with other prekindergarten assessments in order to develop item banks that measure children's cognitive academic language proficiency; and (2) examines the extant literature on differences in skill attainment of preschool male and female children and children from ethnic minority groups identified as LEP or ELL (English language Learners). The paper also addresses some of the pitfalls to be aware of in assessing ethnically and socioeconomically diverse populations of preschool children, assessment items and item wording that might present difficulties for ELL children, and issues involved in using dichotomous or polytomous rating scales for scoring performance assessments. (Contains 47 references, and 3 tables and 2 figures of data. An appendix contains a Code of Fair Testing Practices in Education.) (Author/RS)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 459 473

Running Head: Early childhood cognitive assessments

A Review of the Literature on Early Childhood Cognitive Academic Language Assessments Suitable for Use in Ethnically Diverse School Systems

Marie Miller-Whitehead
Tennessee Valley Educators for Excellence
TVEE.ORG

Paper Presented
At the Annual Meeting
Of the
Mid-South Educational Research Association
Little Rock, AR
November 14, 2001

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY HAS BEEN GRANTED BY

M. Miller-Whitehead

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

For additional information contact the author at:

Marie Miller-Whitehead
PO Box 2882
Muscle Shoals, AL 35662
marie@tvee.org

BEST COPY AVAILABLE

2A

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

CS 217 779

A Review of the Literature on Early Childhood Cognitive Academic Language Assessments Suitable for Use in Ethnically Diverse School Systems

Abstract

Although there are many regions of the country that have historically been ethnically and linguistically diverse, school system data as well as information from the 2000 census and other sources has indicated that many hitherto homogeneous regions of the country and their educational institutions have experienced significant growth in populations of immigrant children and their families. Thus many school systems have faced new challenges to meet the needs of these children from diverse cultural and linguistic backgrounds, including issues related to providing bilingual or ESL education for their Limited English Proficient children.

Sources were selected from current extant psychometric reviews, the U.S. Department of Education, the Council of Chief State School Officers, the Center for Applied Linguistics, the various ERIC Clearinghouses, the U.S. Census Bureau, and position and policy papers on ESL and ELL student testing and placement from AERA, CRESST, and CREDE.

The review examined the literature for approaches used to determine content, construct, and predictive validity of kindergarten cognitive diagnostic assessments and their usefulness for making decisions about prekindergarten students and groups of students from ethnically diverse populations. The paper sought to identify most current literature related to examination of item domain goals and objectives, sources for sample items that could be useful to equate with other prekindergarten assessments for the purpose of developing item banks that measure children's cognitive academic language proficiency, and to examine the extant literature on differences in skill attainment of preschool male and female children and children from ethnic minority groups identified as LEP or ELL.

Also addressed were some of the pitfalls to be aware of in assessing ethnically and socioeconomically diverse populations of preschool children, assessment items and item wording that might present difficulties for ELL children, and issues involved in using dichotomous or polytomous rating scales for scoring performance assessments.

A Review of the Literature on Early Childhood Cognitive Academic Language Assessments Suitable for Use in Ethnically Diverse School Systems

The impetus for this paper arose from two interests: first, an examination conducted by the author of State Education Report Cards and school district accountability indicators, and second, research conducted by the author during time spent developing, administering, scoring, analyzing, and reporting student achievement in an urban school district. The district used the state mandated SAT9 to measure student progress in grades 3, 5, and 8 but each year developed, administered, and scored its own criterion-referenced language and math assessments aligned to district and state curricular objectives (Standards of Learning) for children in grades pre-K, K, 1, 2, 4, and 7 and also designed and disseminated reports of the results of these tests to parents. The district had a mobile population, 47% qualified for free and reduced price meals, and a large percentage of Hispanic and other language minority students. Thus there was a frequent need to assess students in several areas including proficiency in English and level of academic preparation, as students could enter the school system with backgrounds ranging from little or no formal schooling and little or no proficiency in English to excellent academic preparation and a high level of proficiency in English (Table 1).

Insert Table 1 about here

Issues in Testing Ethnic and Language Diverse (LEP) Children

It is important to recognize that a child may be limited English proficient with little formal schooling, i.e., low functioning in many school readiness indicators, and yet be of normal or above normal intelligence. Thus, care should be taken in placement of these students so that they are not inappropriately placed in special education classes or otherwise misidentified (Canino & Spurlock, 1994; Dana, 1993; Rogers, 1998; Scott, Duel, Jean-François, & Urbano, 1996). Although some of the assessments included in this paper have been equated with scores on intelligence tests they should be used with other measures to determine proper placement and intervention for LEP students whose families have limited formal schooling. The case of Diana v California Board of Education is often cited in situations concerning the administration of IQ tests in English to children who are not native speakers of English (Camilli & Shepard, 1994). It is now generally agreed that a measure of receptive or productive vocabulary is not necessarily a measure of intelligence and many tests such as the Peabody Picture Vocabulary Test (PPVT) that were once considered to be intelligence tests are now classified as measures of vocabulary knowledge or cognitive academic language proficiency (Camilli & Shepard). Cognitive academic language proficiency (CALP) is a term that has been defined as proficiency in using language that (a) provides few contextual clues to meaning and (b) is about topics that are not easy and require thought (Cummins, 1981,1989). Immigrant children generally can acquire conversational proficiency in two years while it may take from five to seven years for ELL, ESL

or LEP children to attain levels of cognitive academic language proficiency approaching that of their native English speaking peers. With bilingual children it is also often necessary for an assessment to be able to determine the child's dominant language, as a non-native speaker may be bilingual in some cognitive areas but not in others.

The language testing community has long recognized that a necessary preliminary to providing appropriate instruction for limited English proficient (LEP) children involves assessment of both their English proficiency and their readiness for school or prior exposure to learning experiences in their own language. Many children who are non-native speakers of English may have had opportunities to interact with native English speaking peers and to acquire an informal vocabulary, more commonly referred to as basic interpersonal communication skills (BICS) but may have had little or no opportunity to acquire cognitive academic language proficiency (Cummins, 1981, 1989; Rong & Preissle, 1998). Thus, their range of vocabulary as well as their pronunciation has been influenced by their peer associations, by their adult care-givers, and by their native language. According to Rong and Preissle (1998), while many immigrant children come to school ahead of their U.S. peers in some subject areas, "students from rural areas of Asia, Africa, and Central America may arrive at elementary school not knowing how to use a pencil or eraser, virtually illiterate, and unable to perform basic computation in their native languages" (p. 145). While immigrant intake centers that serve multiple language populations do less than perfect jobs in assessment of children's language and skill abilities for placement purposes, the average classroom teacher does even worse in assessing proper placement and in evaluating and predicting future reading achievement (Flynn & Rahbar, 1998). U. S. Census data from 1990 indicated that as many as 20% of young children who had lived in the U. S. for three years or less did not speak any English and that in the same census, 40% of all foreign born children reported that they had minor to moderate problems with English, were severely limited, or spoke no English at all (Rong & Preissle). According to the U. S. Department of Education, on the average approximately 9% of students in high percentage minority schools are LEP, while only 1% of students in schools with low percentages of minorities are LEP (Han, Baker, & Rodriguez, 1997).

It is therefore quite important to be able to identify the academic language preparation level of children from very diverse socioeconomic backgrounds and also (for the language minority students) to determine their primary or dominant language so that they can receive appropriate instructional support, such as early intervention for a learning disability or ESL services or both. The identification of learning disabilities in limited English proficient children whose families may have limited formal schooling can present special challenges to classroom teachers and counselors. In districts with high percentages of Title I qualified children, instruments should be capable of measuring progress on skills mandated under the requirements of the Title I program, which mandates that all Title 1 children be included in assessments and that data on student achievement be disaggregated at the State, district, and school level by ethnicity and race, gender, by English proficiency status, migrant or non-migrant status, disabled or non-disabled, and economically disadvantaged compared to non-economically disadvantaged. According to a report prepared by the Council of Chief State School Officers (1992), "A student's profile should contain information about his or her linguistic and general cognitive abilities in the native language" and "content testing and assessment in the student's native

language should be part of the overall assessment strategy." The same report provides guidelines for the identification of a student as FEP or fully English proficient.

President Bush has supported enactment of a broad federal education initiative to support achievement of all children under the No Child Left Behind Act of 2001. Federal legislation currently being considered such as S. 303 (Three R's Act) cites "lack of appropriate performance and assessment standards that distinguish between language ability and academic achievement so that State educational agencies and local educational agencies are equally as accountable for the achievement of limited English proficient students in academic content while the students are acquiring English language skills as the agencies are for enabling the students to acquire those skills." The same proposed legislation additionally states that priority should be given to "expanding the range of accommodations available to limited English proficient students and students with disabilities to improve rates of inclusion and to include instructional material development and modified assessment practices that are culturally and ability appropriate, respectively." H.R. 1513 provisions before the 107th Congress include the following language regarding high stakes testing of LEP students:

> the State educational agency or local educational agency provides appropriate accommodations and alternative assessments for students with limited English proficiency (if the agency involved determines that the students have not achieved sufficient English proficiency to ensure that the test will validly and reliably measure the subject matter knowledge and skills of the students), including--
> (A) the use of a test other than an English -only test;
> (B) the use of alternate assessments (consisting of psychometrically equivalent tests in the students' native language) in order to provide such students with a valid and reliable opportunity to demonstrate what the students know and can do; and
> (C) in a case in which the Secretary of Education determines that more than 5 percent of the students enrolled in kindergarten through grade 12 in a State are members of a single language minority group and are limited English proficient--
> (i) the assessment of the students in that group using tests developed in the language of that group, if the State or
> local educational agency determines that such tests are more likely than English -only tests to yield accurate and reliable information regarding what those students know and can do; or
> (ii) if the language of the group is oral or unwritten or, in the case of Alaska Natives and other American Indians, if the predominant language of the group is historically unwritten, the furnishing of oral instructions, assistance, and other necessary information to such students relating to the English -only test; and
> (11) the test is not used for a decision about promotion or placement in special education for a child below the age of 8 or third grade. (H.R. 1513, April 4, 2001)

In addition to the federal recommendations cited above, many states and local education agencies have specific guidelines to address testing accommodations and exclusion or inclusion of ELL or LEP students in testing programs, and 17 states require students to pass a high school

graduation test to receive a regular diploma, with only one state (Texas) allowing students to take a Spanish language version of the high school exit exam (Alcala, 2000; Rivera, Vincent, Hafner, & LaCelle-Peterson, 1997). Recommendations for the specific testing conditions or accommodations that may be provided for students whose native language is other than English are provided in Table 2 (Shepard, Taylor, & Betebenner, 1998).

Insert Table 2 about here

Whether or not and how to provide an ESL or bilingual program within a school district is often determined by numbers of students who would be eligible to participate, with some states and districts establishing a cutoff point of 8 students, and others 20 students per the *Lau* recommendation. While scores on formal assessments are not the only criteria for a child to be considered for a bilingual or ESL program, tests that are often used to determine eligibility are student scores on the Oral Proficiency Test, the California Achievement Test (CAT), the Language Assessment Scales (LAS), the Bilingual Syntax Measure (BSM), and the Iowa Test of Basic Skills (Sevilla, 1992). Cummins (1999) posits that the existence of bilingual programs should not rest solely on the results of the imperfect measures that are used to assess and place students, arguing that there is value in providing students with educational opportunities in their native languages.

Issues in Alignment of Early Childhood Assessments with Cognitive Academic Language Objectives

Most professional organizations and many individuals in the field of test and assessment development and analysis have published their own guidelines for the selection of instruments suitable for the assessment of students (Rudner, 1993, 1994)

Childhood assessments have been developed to measure every type of trait from emotional maturity, intelligence, motor skills, visual acuity, hearing, self-esteem, to acculturation. The criteria for this review of literature on prekindergarten academic language assessment instruments was as follows:

1. The instrument should be suitable for the assessment of skills considered to measure either development or achievement (such as vocabulary that preschool children would be expected to know in kindergarten) rather than strictly behavior, ability, or intelligence.
2. A Spanish language version of the instrument (and other languages if available) should be available from the publisher or author.
3. Administrator's manuals and scoring scales or rubrics with norm group data for each language version should be available from the author or publisher.
4. Preference was given to including instruments that had been reviewed or that had been cited in published research studies.

It is perhaps appropriate to examine briefly the kinds of knowledge that preschool and kindergarten children are generally expected to be familiar with, that is, curricular goals or objectives of the preschool or kindergarten year. La Marca (2001) identifies two broad areas of

match that should be present between curricular objectives and assessment, these identified as (a) content (broad area, range, and balance) and (b) depth. For example, the NCES-administered Early Childhood Longitudinal Study of Kindergarten children (ECLS-K) includes assessments for student attainment in four broad content areas: (a) cognitive skills, (b) social skills, (c) physical well-being, and (d) approaches to learning (West, Denton, & Germino-Hausken, 2000). In addition to items developed specifically for the ECLS-K, the assessment includes items from the CCB, the PIAT, the PPVT-R, the PTCS, and the WJ-R. However, the ECLS-K is administered only in English and according to NCES data, children whose scores on the Oral Language Development Scale (OLDS) indicated they were not proficient in English were not included in the ECLS-K sample population (West, Denton, & Reaney). The cognitive skill areas assessed on the ECLS-K are general knowledge, literacy skills, and math skills. The specific literacy content skills that are assessed are letter recognition, beginning sounds, ending sounds, sight words and words in context. Mathematics skills are numbers-shapes, relative size, ordinality, addition-subtraction, and multiplication-division. Figures 1 and 2 provide information on children's achievement in reading and mathematics skills from the beginning of their kindergarten year in fall of 1998 to the end of the year in spring of 1999. Not surprisingly, there is extensive research showing that knowledge of beginning and ending sounds and of upper and lower case letters of the alphabet are strong predictors of reading success, although some studies indicate that phonological awareness is a better predictor of superior reading than of reading problems (Churchill, Durdel, & Kenney, 1998; Engel, 1991; Kirby & Parrila, 1999; Miller-Whitehead, 2001; Snow, Burns, & Griffin, 1998; Zill, Collins, West, & Germino-Hausken, 1995).

However, the previously mentioned skill areas of letter recognition, beginning and ending sounds, sight words and words in context are descriptors only; they are not indicators of depth of knowledge nor are they presented in the form of measurable objectives. Depth, of course, refers to the variety of ways and the detail in which the content area can be demonstrated: can the child recognize the letters of the alphabet when presented out of sequence, in an unfamiliar typeface, or in isolation from contextual organizers? Can the child follow words on a page, pointing to a word as the teacher pronounces it, scan from left to right (and back to the left if there are two lines of text on the page) and top to bottom of a page, recognize beginning consonant sounds, match voice to print, recognize alliteration, segmentation, rhyme, or otherwise demonstrate phonemic awareness and differentiate between minimal pairs? The next step in the alignment of standards to test is for each goal or criteria of the assessment to be tied to one or more measurable objectives. More measurable objectives for each curricular standard provides more depth to the assessment system. Thus, a measurable objective for alphabet recognition might be stated, "correctly names at least 23 of 26 upper case letters of the alphabet presented out of sequence." While it is not recommended that IQ tests or other measures of intelligence include items that require young children or LEP immigrants to identify coins or count money (many have never had or even seen American money), it is certainly appropriate to include this skill as a measurable achievement objective for English vocabulary ("point to the nickel, dime, quarter, penny") as well as for counting in math ("point to the coins that add up to 20 cents").

The issue of testing young children, and ages at which it is appropriate to do so, varies widely according to state or district policy and recommendations of early childhood

professionals (Kamii, 1990). As of 2000-2001, some states that have mandated statewide accountability tests do not begin statewide testing of students until grade 2 or grade 3; Georgia requires that all kindergarten students, unless exempted by IEP or IAP, be assessed with components of the Georgia Kindergarten Assessment Program-Revised (GKAP-R) in the fall, winter, and spring of the kindergarten year and their progress monitored through grade 1.

Selected Assessments Appropriate for Assessment of Literacy of Ethnic and Language Diverse Preschool and Kindergarten Children

A search of Internet and print resources showed that there were an enormous number of early childhood assessments available both in and out of print, many described in the various editions of Tests in Print and the Mental Measurements Yearbooks published by the Buros Institute or listed in the ETS Test Collection Catalog (1986). Online resources are the ETS database collection including Tests in Microfiche, the database abstracts and full text digests of the ERIC Clearinghouse on Assessment and Evaluation, and assessments abstracted at the Center for Excellence and Equity in Education. Some of these were well-known, popular, and widely used assessments developed by the larger testing organizations and companies while others were quite specialized or had been in print for only a short time. Several general comments are perhaps appropriate as an overview to this section of the paper. First and most importantly, this listing does not presume to be a comprehensive guide to early childhood assessments, either criterion-referenced or norm-referenced, and many excellent assessments of each type were not included. Since the selection criteria specified assessments of school readiness such as literacy and vocabulary skills rather than ability or IQ, tests such as the WISC-III, the WPPSI, the Slossen, and the Stanford-Binet were not included even though they generate verbal or vocabulary ability scores. The Raven's Progressive Matrices test and the Test of Non-Verbal Intelligence-2 (TONI-2) are nonverbal tests of ability rather than measures of cognitive academic language and thus were not included. Other instruments were not included because they were not available in a Spanish language version from the publisher. These included the ITBS, the SESAT, the CogAT, and the Metropolitan Achievement Test-7 (MAT-7) although Spanish language informational materials for parents were available from the publisher for the MAT-7 (Drummond, 2000; Tests in Print IV, 1994).

With few exceptions published test reviews were not overwhelmingly favorable, with the reviewers providing a variety of reasons for their opinions. In general reviewers found that the authors or publishers of many of the assessments, even some of the most widely used in education, did not adequately describe the population for which the assessment was intended or did not provide adequate data about the pilot population. Many of the assessments that did provide means, standard deviations, and SEM for the total population did not provide disaggregated data by gender, ethnicity, English proficiency, SES, handicapping condition, or race. In some instances, where disaggregated norm data were provided for English versions of a test they were not provided for the Spanish version. Other limitations had to do with available scores or scales and with manuals and directions for scoring, particularly for tests in Spanish or that are used to assess level of English proficiency for LEP students, with some reviewers feeling that scoring rubrics did not adequately describe measurable objectives or the level of achievement necessary to demonstrate attainment of the objective, for example, "the student is

able to pronounce the indicated word without being made fun of by classmates." For some tests documentation, and thus mechanisms for assuring standardization, were minimal. However, many of these limitations have been addressed through the work of independent researchers who have used some of these instruments in hundreds of published research studies, thus providing a basis for comparison of populations of varying levels of ability under a variety of treatment conditions. In fact, there is extensive research that equates scores on many of the early childhood literacy or skills tests with scores on some of the IQ tests mentioned above.

Table 3 provides a selected list of prekindergarten and kindergarten assessments that are available in both English and Spanish for the identification of literacy (cognitive academic language proficiency), number skills and language dominance.

Insert Table 3 about here

Discussion

The Brigance Diagnostic Comprehensive Inventory of Basic Skills (CIBS) is a criterion-referenced test used extensively in special education for documentation of IEP, although one reviewer describes it as norm referenced in an examination of its standardization procedures, reliability, and validity (Glascoe, 1999). It is available in two forms as well as a Spanish and Portuguese version. According to the reviewers, the manuals supply little information on how cut scores were established when they are provided, no standards for mastery, and no data on reliability, validity, or item analysis (Mills, 1985; Smith, 1985; Swerdlik, 1985; Witt, 1985). Each response can only be coded for mastery or nonmastery (a gross measure according to one reviewer), and scores are reported as grade placement scores rather than as attainment to objectives, thus perhaps not providing sufficient data for results to be useful as a tool to improve teaching and learning. The assessment is commonly used with bilingual, ESL, and migrant students to determine language dominance. The assessment yields a score for (a) readiness; (b) speech; (c) word recognition; (d) oral reading; (e) reading comprehension; (f) word analysis; (g) listening; (h) writing and alphabetizing; (i) numbers and computation; and (j) measurement (Tests in Print IV, 1994).

The Kaufman Assessment Battery for Children (K-ABC) has perhaps been cited in more published research than almost any other early childhood instrument. It is most commonly used as an intelligence or ability measure and yields 16 subtest scores, several of which are achievement subtests of vocabulary and arithmetic skills; it is available in languages other than English. Another instrument, the Kaufman Survey of Early Academic and Language Skills (K-SEALS) yields 8 scores that are perhaps more aligned with the skills objectives of the other assessments listed here than is the K-ABC, but the publisher does not indicate that the K-SEALS is available in Spanish. Scores on the K-ABC have been correlated with scores on the WISC-R, the Stanford-Binet, the PPVT-R, and the WJ-R, and the instrument has been used in published research studies of gender, giftedness, ethnicity and SES (Mexican-American, native American, and Head Start children), learning disabled and normal children, and LEP populations. It

includes a non-verbal scale for assessment of non-English-speaking children (Tests in Print IV, 1994).

The Language Assessment Scales Oral (LAS-O) is, as the name implies, an oral assessment that is recommended for use in grades 1-12, rather than for pre-K and kindergarten children. However, it is mentioned here as it is available in both an English and Spanish version with scores for (a) school vocabulary picture naming; (b) listening comprehension (c) story retelling; and (d) minimal sound pairs and phonemes. Reliability for the minimal pairs, phonemes and vocabulary scores is reported to be in the high .80s with listening score reliabilities ranging from .30s to the .90s (Tests in Print IV, 1994).

The Lollipop-Revised has been in print since 1981 and yields 5 scores: (a) colors, shapes, copying shapes; (b) picture description and spatial recognition; (c) identification of numbers and counting; (d) identification of letters and writing; and (e) total. There is a Spanish version and the test has been used to determine language dominance of bilingual children. Correlations of scores on the MRT and the Lollipop have ranged from .71 to .86 for total test scores (Johnson, 1992; Oosterhof, 1992). Norms and some data are provided, including gender and ethnicity of the sample population, and although factor scores were obtained item loadings were not reported. In later published research, the author conducted a Rasch analysis to examine equivalence of the Spanish and English versions of the Lollipop test with populations of bilingual, rural, disadvantaged, White, Black, and Hispanic students, finding that there was no evidence of bias on the total test (Lang, Chew, & Vargas, 1993).

The original of the Peabody Picture Vocabulary Test (PPVT) has been around since 1959, and the PPVT-R since 1981. A total score is reported and a Spanish version is available. It is a test of vocabulary comprehension and requires that the examinee demonstrate understanding of a word spoken by the examiner by pointing to one of four pictures. Research has been conducted with a computer-adapted version (Maguire, 1991). Although originally it was considered to be a test of ability, more recently it has been used as a measure of cognitive language achievement. PPVT-R scores correlate to both achievement tests (WRAT) and ability tests such as the WISC-R and the K-ABC. As one of the most often referenced of all early childhood assessments, it has been the subject of extensive review and used as a measure of achievement in research on gender, gender identity, learning disabilities, limited English proficiency, behavioral problems, mnemonics, ethnic minorities, native American children, "handedness," Headstart programs, and bilingualism, to name only a few (Maddux, 1999; Smith, 1991; Tests in Print IV, 1994; Umbel, 1992; Vance & Stone, 1990).

Raw scores on the Woodcock-Muñoz Language Survey can be entered into a software program that will compute an individual's CALP (cognitive academic language proficiency) and the levels of communication which the individual would find easy or difficult. The test is recommended for use to (a) determine eligibility for bilingual services; (b) classify students by language ability for instruction; (c) assess progress in English-only instruction; (d) evaluate ESL programs; (e) determine language dominance of bilingual Spanish-English speakers; and (f) assess language proficiency of subjects in research studies. The test is said by its publishers to be

appropriate for ages 2 to 90 and consists of four parts, (a) picture vocabulary; (b) verbal analogies; (c) letter-word identification; and (d) dictation; these four scores can be combined to yield three cluster scores so that a total of seven scores are available. Rasch modeling was used to equate the English and Spanish versions of the assessment and W scores are reported, although reviewers did not feel that the equating procedure was adequately described in terms of group scores that were compared or how test items in different languages were compared, one reviewer noting that this was the "Achilles heel" of the equating system. Reliabilities range from .80s to .90s for each of the four scores for the English version but are not reported for the Spanish version (Crocker, 1998; Kao, 1998).

Conclusions

Even though the author has had considerable experience in language testing, particularly in foreign language, this project turned out to be more challenging than anticipated due to the sheer numbers of instruments that were available even when limited to assessments of language proficiency. This list is not comprehensive and it is certain that many excellent assessments that fit the criteria have not been included. The original purpose of this paper was to identify instruments that could be used to assess pre-kindergarten and kindergarten children's cognitive academic language proficiency in either English or Spanish and also to determine the dominant language of bilingual children. Each of the assessment instruments described in this paper has been used for these purposes and although reviewers had differing opinions about issues of validity, reliability, and standardization for some of the assessments there has been extensive published research conducted with all of them, much of it by highly-regarded measurement and testing professionals. Although this may be construed as preaching to the choir, those who may be involved in selection of an appropriate pre-K assessment instrument for multicultural groups of children are encouraged to review the Code of Fair Testing Practices in Appendix A.

References

Alcala, A. (2000). A framework for developing an effective instructional program for limited English proficient students with limited formal schooling. Practical Assessment, Research & Evaluation, 7(9). Available online: http://ericae.net/pare/getvn.asp?v=7&n=9.

Camilli, L., & Shepard, G. (1994). Methods for identifying biased test items. Richard M. Jaeger (Ed.). Measurement Methods for the Social Sciences (Vol. 4). Thousand Oaks, CA: Sage Publications.

Canino, I. A., & Spurlock, J. (1994). Culturally diverse children and adolescents: Assessment, diagnosis, and treatment. New York: Guilford.

Churchill, K., Durdel, J., & Kenney, M. (1998). Hear it, feel it, see it: Improving early reading acquisition through a multisensory phonemic awareness approach. (ERIC Document Reproduction Service No. ED 420 049)

Council of Chief State School Officers (1992). Summary of recommendations and policy implications for improving the assessment and monitoring of students with Limited English Proficiency. Continuation of Summary of State Practices Concerning the Assessment of and Data Collection about Limited English Proficiency Students. Washington, DC: Author.

Crocker, L. (1998). Review of the Woodcock-Muñoz Language Survey. In J. C. Impara & B. S. Plake (Eds.), The thirteenth mental measurements yearbook (pp. 1143-1145). The Buros Institute of Mental Measurements. Lincoln, NE: The University of Nebraska Press.

Cummins, J. (1999). Alternative paradigms in bilingual education research: Does theory have a place? Educational Researcher, 28(7), 26-32.

Cummins, J. (1989). Empowering minority students. Sacramento, CA: California Association for Bilingual Education.

Cummins, J. (1981). Schooling and language minority students: A theoretical framework. Los Angeles: Evaluation, Dissemination, and Assessment Center, California State University.

Dana, R. H. (1993). Multicultural assessment perspectives for professional psychology. Boston: Allyn & Bacon.

Drummond, R. J. (2000). Appraisal procedures for counselors and helping professionals (4th ed.). Upper Saddle River, NJ: Prentice-Hall.

Engel, B. S. (1991). An approach to assessment in early literacy. In C. Kamii, M. Manning, & G. Manning (Eds.), Early literacy: A constructivist foundation for whole language (p. 128). NEA Early Childhood Education Series. Washington, DC: National Education Association of the United States.

The ETS test collection catalog: Achievement tests and measurement devices (Vol. 1). (1986). Compiled by Test Collection of Educational Testing Service. Phoenix: The Oryx Press.

Flynn, J. M., & Rahbar, M. H. (1998). Improving teacher prediction of children at risk for reading failure. Psychology in the Schools, 35(2), 163-172.

Glascoe, F. (1999). Brigance Comprehensive Inventory of Basic Skills-Revised. Diagnostique, 24(1-4), 41-51.

Han, M., Baker, D., & Rodriguez, C. (1997). A profile of policies and practices for Limited English Proficient students: Screening methods, program support, and teacher training. [SASS 1993-1994]. NCES 97-472. Project Officer, Peggy Quinn. Washington, DC: U.S. Department of Education. National Center for Education Statistics.

Johnson, S. (1992). Review of the Lollipop Test: A diagnostic screening test of school readiness-revised. In J. J. Kramer & J. C. Conoley (Eds.), The eleventh mental measurements yearbook (pp. 476-478). The Buros Institute of Mental Measurements. Lincoln, NE: The University of Nebraska Press.

Kamii, C. (Ed.). (1990). Achievement testing in the early grades: The games grownups play. Washington, DC: National Association for the Education of Young Children.

Kao, C. (1998). Review of the Woodcock-Munoz Language Survey. In J. C. Impara & B. S. Plake (Eds.), The thirteenth mental measurements yearbook (pp. 1143-1145). The Buros Institute of Mental Measurements. Lincoln, NE: The University of Nebraska Press.

Kirby, J. R., & Parrila, R. K. (1999). Theory-based prediction of early reading. Alberta Journal of Educational Research, 45(4), 428-447.

Lang, W. S., Chew, A. L., & Vargas, C. (1993). Rasch model applications to determine the equivalence of a readiness test in two languages. Paper presented at the International Objective Measurement Workshop Emory University (Atlanta, GA, April 10, 1993). (ERIC Document Reproduction Service No. ED 379 293)

Maddux, C. D. (1999). Peabody Picture Vocabulary Test III (PPVT-III). Diagnostique, 24(1-4), 221-228.

Maguire, K. (1991). Computer-adapted PPVT-R: A comparison between standard and modified versions within a elementary school population. Psychology in the Schools, 28(3), 199-205.

La Marca, Paul M. (2001). Alignment of standards and assessments as an accountability criterion. Practical Assessment, Research & Evaluation, 7(21). Available online: http://ericae.net/pare/getvn.asp?v=7&n=21

Lau v. Nichols, 414 U.S. 563 (1974).

Miller-Whitehead, M. (2001). Evaluation of a Kindergarten Diagnostic Assessment by gender and ethnicity. Paper presented at the Annual Meeting of the American Educational Research Association (Seattle, WA, April 2001). (ERIC Document Reproduction Service No. ED 453 250)

Mills, C. N. (1985). Review of Brigance Comprehensive Inventory of Basic Skills. In J. V. Mitchell, Jr. (Ed.), The ninth mental measurements yearbook (Vol. 1), (pp. 213-214). The Buros Institute of Mental Measurements. Lincoln, NE: The University of Nebraska Press.

Murphy, L., Conoley, J., & Impara, J. (1994). Tests in print IV (Vols. 1-2). The Buros Institute of Mental Measurements. Lincoln, NE: The University of Nebraska Press.

Oosterhof, A. C. (1992). Review of the Lollipop Test: A diagnostic screening tool of school readiness-revised. In J. J. Kramer & J. C. Conoley (Eds.), The eleventh mental measurements yearbook (pp. 478-479). The Buros Institute of Mental Measurements. Lincoln, NE: The University of Nebraska Press.

Rivera, C., Vincent, C., Hafner, A., & LaCelle-Peterson, M. (1997). Statewide assessment programs: policies and practices for the inclusion of limited English proficient students. Practical Assessment, Research & Evaluation, 5(13). Available online: http://ericae.net/pare/getvn.asp?v=5&n=13.

Rogers, M. R. (1998). Assessment of culturally linguistically diverse children and youth. In H. B. Vance (Ed.), Psychological assessment of children (pp. 358-384). New York: Wiley.

Rong, X. L., & Preissle, J. (1998). Educating immigrant students: What we need to know to meet the challenges. Thousand Oaks, CA: Corwin Press.

Rudner, L. (1994). Questions to ask when evaluating tests. Practical Assessment, Research, & Evaluation, 4(2). Available online: http://ericae.net/pare/getvn.asp?v=4&n=2.

Rudner, L. (1993). Test evaluation. Available online at: http://ericae.net.

Scott, M. S., Duel, L. S., Jean-François, B., & Urbano, R. C. (1996). Identifying cognitively gifted ethnic minority children. Gifted Children Quarterly, 40(3), 147-153.

Sevilla, J. (1992). Bilingual education: The last 25 years. In P. Anthony and S. L. Jacobson (Eds.), Helping at-risk students: What are the educational and financial costs? (pp. 38-65). Thirteenth annual yearbook of the American Educational Finance Association. Newbury Park, CA: Corwin Press.

Shepard, L., Taylor, G., & Betebenner, D. (1998). Inclusion of Limited-English-Proficient Students in Rhode Island's Grade 4 Mathematics Performance Assessment. CSE Technical Report 486. CRESST/University of Colorado at Boulder. Santa Cruz, CA: Center for Research on Evaluation, Diversity, and Excellence (CREDE).

Smith, T. C. (1991). Relationship between the Peabody Picture Vocabulary Test-Revised, Wide Range Achievement Test-Revised, and Wechsler Intelligence Scale for Children-Revised. Journal of School Psychology, 29(1), 53-56.

Snow, C., Burns, S., & Griffin, P. (1998). Predictors of success and failure in reading. In C. Snow, S. Burns, & P. Griffin (Eds.), Preventing reading difficulties in young children (pp. 100-133). Washington, DC: National Academy Press.

Swerdlik, M. E. (1985). Review of Brigance Comprehensive Inventory of Basic Skills. In J. V. Mitchell, Jr. (Ed.), The ninth mental measurements yearbook (Vol. 1), (pp. 214-217). The Buros Institute of Mental Measurements. Lincoln, NE: The University of Nebraska Press.

Umbel, V. M. (1992). Measuring bilingual children's receptive vocabularies. Child Development, 63(4), 1012-1020.

Vance, H., & Stone, J. (1990). Peabody Picture Vocabulary Test-Revised (PPVT-R). Diagnostique, 15(1), 149-160.

West, J., Denton, K., & Germino-Hausken, E. (2000). America's kindergartners. Early Childhood Longitudinal Study, Kindergarten Class of 1998-1999, Fall of 1998. Statistical Analysis Report NCES 2000-170, National Center for Education Statistics. Washington, DC: United States Department of Education, Office of Educational Research and Improvement.

West, J., Denton, K., & Reaney, L. (2000). The kindergarten year. Findings from the Early Childhood Longitudinal Study, Class of 1998-1999. National Center for Education Statistics Washington, DC: United States Department of Education.

Witt, J. C. (1985). Review of Brigance Comprehensive Inventory of Basic Skills. In J. V. Mitchell, Jr. (Ed.), The ninth mental measurements yearbook (Vol. 1), (pp. 217-218). The

Buros Institute of Mental Measurements. Lincoln, NE: The University of Nebraska Press.

Zill, N., Collins, M., West, J., & Germino-Hausken, E. (1995). School readiness and children's developmental status. ERIC Digest. (ERIC Document Reproduction Service No. ED 389 475)

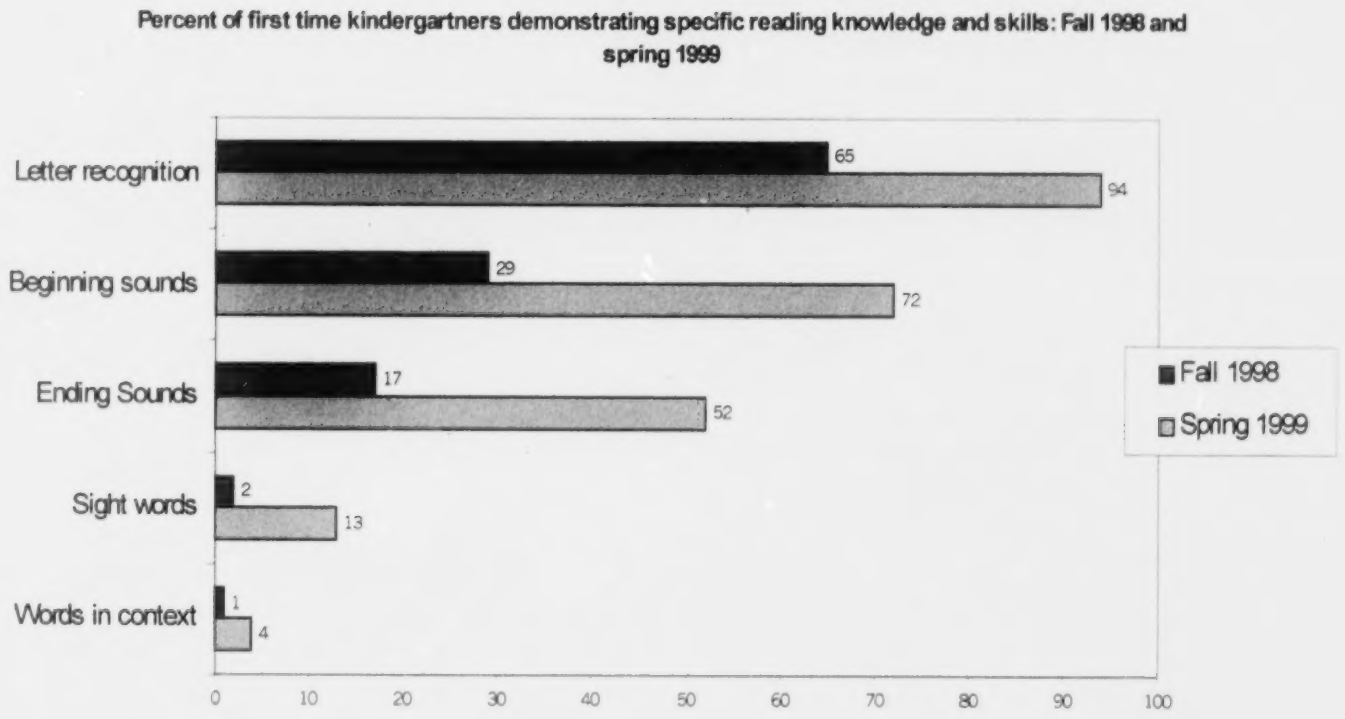

Figure 1. Language skills attainment on the ECLS-K of kindergarten cohort fall 1998 and spring 1999

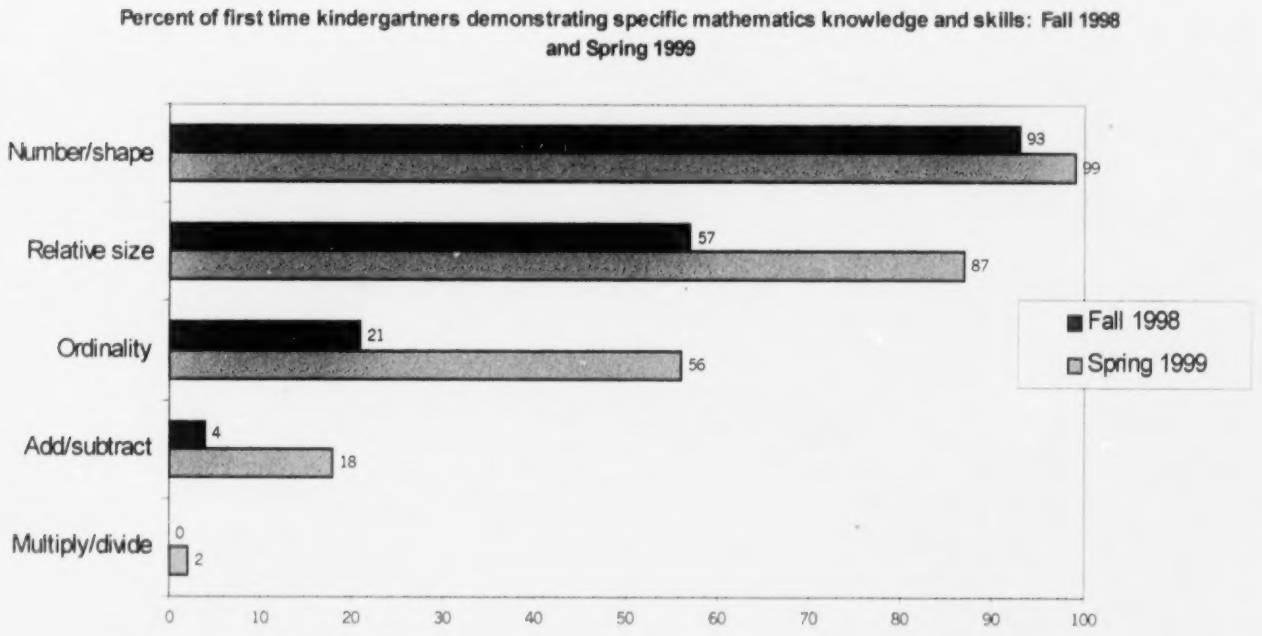

Figure 2. Math skills attainment on ECLS-K of kindergarten cohort in fall of 1998 and spring of 1999.

Table 1

Academic and Language Proficiency Levels of Non-Native English Speakers

School Readiness Level	Academic Preparation	English Proficiency
1 LFS/LEP	Low (LFS)	Low (LEP)
2 LFS/BICS Bilingual	Low (LFS)	High (BICS Bilingual)
3 EFS/LEP	Extensive (EFS)	Low (LEP)
4 EFS/CALP Bilingual	Extensive (EFS)	High (CALP Bilingual)

Note. LFS is an acronym used to denote children with limited formal schooling. BICS and CALP are acronyms for basic interpersonal communications skills and cognitive academic language proficiency, respectively. The ETS and ACTFL scoring rubric uses a 6 level proficiency scale with scores ranging from 0 to 5, with 5 as the highest level of language proficiency, equivalent to that of a well-educated native speaker.

Table 2

Two Categories of Accommodations for English Language Learners

Test	Test Procedure
Assess in the native language	Extra assessment time
Text changes in vocabulary	Breaks during testing
Modification of linguistic complexity	Administration in several sessions
Addition of visual supports	Oral directions in the native language
Use of glossaries in native language	Small-group administration
Use of glossaries in English	Separate room administration
Linguistic modification of test directions	Use of dictionaries
Additional example items	Reading aloud of questions in English
	Answers written directly in test booklet
	Directions read aloud or explained

Table 3

Test	Acronym	Date	Scores	Time	Admin	Language	Author	Publisher
Bilingual Two Language Battery of Tests		1983	4, (phonics, comprehension, oral proficiency)	16-30 min	Group	English, Spanish, Italian, Vietnamese, Portuguese, French	A. Caso	Branden Press
Brigance Diagnostic Assessment of Basic Skills	CIBS (English)	1984	10/24	not given	Indiv	English, Spanish, Portuguese	A. Brigance	Curriculum Associates
Comprehensive Test of Basic Skills, Terra Nova, Supera	CTB	1990	7	117	Group	English Spanish language version is called the Supera	CTB	McGraw-Hill
Iowa Test of Basic Skills/	ITBS	1990	8	125 - 150 min	Group	ITBS, English	Hieronymus, Hoover, Oberley, Cantor, Frisbie, Dunbar, Lewis, Lindquist	Riverside Publishing
La Prueba de Realización		1991	2 (reading & lang, math)	50-70 min	Group	La Prueba de Realizacion, Spanish	N. Cole, E. Trent, D. Wadell	
Kaufman Assessment Battery for Children	K-ABC	1983	16, (6 achievement) *Ability test	50-60 min	Indiv.	English, Spanish and other available	A. Kaufman & N. Kaufman	American Guidance Service
Language Assessment Scales Oral	LAS-O	1987-1991 Sp-1990	6, (vocabulary, minimal pairs, phonemes)	15 min	Indiv	English, Spanish version available	E. DeAvila & S. Duncan	CTBMacmillan /MacGraw-Hill

Table 3 (Cont.)

Early childhood cognitive assessments

Test	Acronym	Date	Scores	Time	Admin	Language	Author	Publisher
Lollipop-Revised		1981-89	5, (colors, shapes, numbers, letters)	15-20 min	Indiv	English, Spanish	A. Chew	Humanics
Metropolitan Readiness Test	MRT	1990	2 Levels, 8, (beginning consonants, sound-letter, language, quantitative)	90 min	Group	English, Spanish	J. Nurss & M. McGauvran	Psychological Corp
Peabody Picture Vocabulary Test-Revised	PPVT-R	1959-1981	total only	10-20 min	Indiv	English, Spanish	L. Dunn, L. Dunn, G. Robertson, & J. Eisenberg	American Guidance Service
Woodcock-Munoz Language Survey	LS-E, LS-S	1993	7, (pictures, letters, oral)	untimed	Indiv	English, Spanish	R. Woodcock & A. Munoz	Riverside

APPENDIX A

CODE OF FAIR TESTING PRACTICES IN EDUCATION
PREPARED BY THE JOINT COMMITTEE ON TESTING PRACTICES

The Code of Fair Testing Practices in Education states the major obligations to test takers of professionals who develop or use educational tests. The Code is meant to apply broadly to the use of tests in education (admissions, educational assessment, educational diagnosis, and student placement). The Code is not designed to cover employment testing, licensure or certification testing, or other types of testing. Although the Code has relevance to many types of educational tests, it is directed primarily at professionally developed tests such as those sold by commercial test publishers or used in formally administered testing programs. The Code is not intended to cover tests made by individual teachers for use in their own classrooms.

The Code addresses the roles of test developers and test users separately. Test users are people who select tests, commission test development services, or make decisions on the basis of test scores. Test developers are people who actually construct tests as well as those who set policies for particular testing programs. The roles may, of course, overlap as when a state education agency commissions test development services, sets policies that control the test development process, and makes decisions on the basis of the test scores.

The Code has been developed by the Joint Committee on Testing Practices, a cooperative effort of several professional organizations, that has as its aim the advancement, in the public interest, of the quality of testing practices. The Joint Committee was initiated by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education. In addition to these three groups the American Association for Counseling and Development/Association for Measurement and Evaluation in Counseling and Development, and the American Speech-Language-Hearing Association are now also sponsors of the Joint Committee.

This is not copyrighted material. Reproduction and dissemination are encouraged. Please cite this document as follows:

Code of Fair Testing Practices in Education. (1988)
Washington, D.C.: Joint Committee on Testing Practices.

(Mailing Address: Joint Committee on Testing Practices,
American Psychological Association, 1200 17th Street, NW,
Washington, D.C. 20036.)

The Code presents standards for educational test developers and users in four areas:

A. Developing/Selecting Tests
B. Interpreting Scores
C. Striving for Fairness
D. Informing Test Takers

BEST COPY AVAILABLE

Organizations, institutions, and individual professionals who endorse the Code commit themselves to safeguarding the rights of test takers by following the principles listed. The Code is intended to be consistent with the relevant parts of the Standards for Educational and Psychological Testing (AERA, APA, NCME, 1985). However, the Code differs from the Standards in both audience and purpose. The Code is meant to be understood by the general public; it is limited to educational tests; and the primary focus is on those issues that affect the proper use of tests. The Code is not meant to add new principles over and above those in the Standards or to change the meaning of the Standards. The goal is rather to represent the spirit of a selected portion of the Standards in a way that is meaningful to test takers and/or their parents or guardians. It is the hope of the Joint Committee that the Code will also be judged to be consistent with existing codes of conduct and standards of other professional groups who use educational tests.

A. DEVELOPING/SELECTING APPROPRIATE TESTS*

Test developers should provide the information that test users need to select appropriate tests.	Test users should select tests that meet the purpose for which they are to be used and that are appropriate for the intended test taking populations.
TEST DEVELOPERS SHOULD:	TEST USERS SHOULD:
1. Define what each test measures and what the test should be used for. Describe the population(s) for which the test is appropriate. appropriate.	1. First define the purpose for testing and the population to be tested. Then, select a test for that purpose and that population based on a thorough review of the available information.
2. Accurately represent the characteristics, usefulness, and limitations of tests for their intended purposes.	2. Investigate potentially useful sources of information, in addition to test scores, to corroborate the information provided by tests.
3. Explain relevant measurement concepts as necessary for clarity at the level of detail that is appropriate for the intended audience(s).	3. Read the materials provided by test developers and avoid using tests for which unclear or incomplete information is provided.
4. Describe the process of test development. Explain how the content and skills to be tested were selected.	4. Become familiar with how and when the test was developed and developed and tried out.
5. Provide evidence that the test meets its intended purpose(s).	5. Read independent evaluations of a test and of possible alternative measures. Look for evidence required to support the claims of test developers.

6. Provide either representative samples or complete copies of test questions, directions, answer sheets, manuals, and score reports to qualified users.	6. Examine specimen sets, disclosed tests or samples of questions, directions, answer sheets, manuals, and score reports before selecting a test.

*Many of the statements in the Code refer to the selection of existing tests. However, in customized testing programs test developers are engaged to construct new tests. In those situations, the test development process should be designed to help ensure that the completed tests will be in compliance with the Code.

TEST DEVELOPERS SHOULD:	TEST USERS SHOULD:
7. Indicate the nature of the evidence obtained concerning the appropriateness of each test for groups of different racial, ethnic, or linguistic backgrounds who are likely to be tested.	7. Ascertain whether the test content and norm group(s) or comparison group(s) are appropriate for the intended test takers.
8. Identify and publish any specialized skills needed to administer each test and to interpret scores correctly.	8. Select and use only those tests for which the skills needed to administer the test and interpret scores correctly are available.

B. INTERPRETING SCORES

Test developers should help users interpret scores correctly.	Test users should interpret scores correctly.
TEST DEVELOPERS SHOULD:	TEST USERS SHOULD:
9. Provide timely and easily understood score reports that describe test performance clearly and accurately. Also, explain the meaning and limitations of reported scores.	9. Obtain information about the scale used for reporting scores, the characteristics of any norms or comparison group(s), and the limitations of the scores.
10. Describe the population(s) represented by any norms or comparison group(s), the dates the data were gathered, and the process used to select the samples of test takers.	10. Interpret scores taking into account any major differences between the norms or comparison groups and the actual test takers. Also take into account any differences in test administration practices or familiarity with the specific questions in the test.
11. Warn users to avoid specific, reasonably anticipated misuses of test scores.	11. Avoid using tests for purposes not specifically recommended by the test developer unless evidence is obtained to

BEST COPY AVAILABLE

	support the intended use.
12. Provide information that will help users follow reasonable procedures for setting passing scores when it is appropriate to use such scores with the test.	12. Explain how any passing scores were set and gather evidence to support the appropriateness of the scores.
13. Provide information that will help users gather evidence to show that the test is meeting its intended purpose(s).	13. Obtain evidence to help show that the test is meeting its intended purpose(s).

C. STRIVING FOR FAIRNESS

Test developers should strive to make tests that are as fair as possible for test takers of different races, gender, ethnic backgrounds, or different handicapping conditions.	Test users should select tests that have been developed in ways that attempt to make them as fair as possible for test takers of different races, gender, ethnic backgrounds, or handicapping conditions.
TEST DEVELOPERS SHOULD:	TEST USERS SHOULD:
14. Review and revise test questions and related materials to avoid potentially insensitive content or language.	14. Evaluate the procedures used by test developers to avoid potentially insensitive content or language.
15. Investigate the performance of test takers of different races, gender, and ethnic backgrounds when samples of sufficient size are available. Enact procedures that help to ensure that differences in performance are related primarily to the skills under assessment rather than to irrelevant factors.	15. Review the performance of test takers of different races, gender, and ethnic backgrounds when samples of sufficient size are available. Evaluate the extent to which performance differences may have been caused of the test.
16. When feasible, make appropriately modified forms of tests or administration procedures available for test takers with handicapping conditions. Warn test users of potential problems in using standard norms with modified tests or administration procedures that result in non-comparable scores.	16. When necessary and feasible, use appropriately modified forms or administration procedures for test takers with handicapping conditions. Interpret standard norms with care in the light of the modifications that were made.

D. INFORMING TEST TAKERS

Under some circumstances, test developers have direct communication with test takers. Under other circumstances, test users communicate directly with test takers. Whichever group communicates directly with test takers should provide the information described below.

TEST DEVELOPERS OR TEST USERS SHOULD:

17. When a test is optional, provide test takers or their parents/guardians with information to help them judge whether the test should be taken, or if an available alternative to the test should be used.

18. Provide test takers the information they need to be familiar with the coverage of the test, the types of question formats, the directions, and appropriate test-taking strategies. Strive to make such information equally available to all test takers.

Under some circumstances, test developers have direct control of tests and test scores. Under other circumstances, test users have such control. Whichever group has direct control of tests and test scores should take the steps described below.

TEST DEVELOPERS OR TEST USERS SHOULD:

19. Provide test takers or their parents/guardians with information about rights test takers may have to obtain copies of tests and completed answer sheets, retake tests, have tests rescored, or cancel scores.

20. Tell test takers or their parents/guardians how long scores will be kept on file and indicate to whom and under what circumstances test scores will or will not be released.

21. Describe the procedures that test takers or their parents/guardians may use to register complaints and have problems resolved.

Note: The membership of the Working Group that developed the Code of Fair Testing Practices in Education and of the Joint Committee on Testing Practices that guided the Working Group was as follows:

Theodore P. Bartell
John R. Bergan
Esther E. Diamond
Richard P. Duran
Lorraine D. Eyde
Raymond D. Fowler

John J. Fremer
(Co-chair, JCTP and Chair, Code Working Group)
Edmund W. Gordon
Jo-Ida C. Hansen
James B. Lingwall

George F. Madaus
(Co-chair, JCTP)
Kevin L. Moreland
Jo-Ellen V. Perez
Robert J. Solomon
John T. Stewart
Carol Kehr Tittle
(Co-chair, JCTP)

Nicholas A. Vacc
Michael J. Zieky

(Debra Boltas and Wayne Camara of the American Psychological Association served as staff liaisons)

Additional copies of the Code may be obtained from the National Council on Measurement in Education, 1230 Seventeenth Street, NW, Washington, D.C. 20036.
Single copies are free.

BEST COPY AVAILABLE